PublicGarbageNet : A Deep Learning Framework for Public Garbage Classification

Ming Zeng, Xiangzhe Lu, Wenkang Xu, Tongxi Zhou, Yinbo Liu
Institute of Robotics and Automation Systems, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, P. R. China
E-mail: zengming@tju.edu.cn

Abstract: Intelligent garbage classification is an important technique for garbage harmless, reducing and recycling treatment. In this paper, we propose a public garbage classification algorithm based on the CNN architecture, namely PublicGarbageNet. The proposed algorithm is a multi-task classification algorithm in which one task identifies four major categories of domestic garbage, and the other task achieves recognition of 10 subclasses garbage. The two classification tasks are related to each other, and the joint loss function is helpful to improve the accuracy of garbage recognition. Considering the fact that the existing garbage datasets are incomplete in classes and small in quantity, we constructed a new pubic garbage dataset including 10 subclasses and a total of 10624 images. In order to obtain better performance, systematically studies such as backbone optimization selection, data augmentation, learning rate optimization, and label smoothing have been made, and finally the accuracy of the optimized model reaches 96.35%.

Key Words: garbage classification, public garbage dataset, multi-task learning, data augmentation

1 Introduction

According to statistics from the China Association of Environmental Protection Industry(CAEPI), about two thirds of China's major cities are surrounded by ever expanding garbage dumps covering area of 500 million square meters, incurring more than 30 billion yuan economic losses a year. These huge volumes of garbage have placed great pressure on the city's waste treatment facilities. Therefore, the research on intelligent classification of household garbage will be helpful to develop advanced domestic garbage treatment technology and equipment.

Object classification is a fundamental task for computer vision in general. The procedure of traditional algorithms for object image classification mainly consists of two steps: feature extraction and classification. Each step has plenty of optional methods (e.g., SIFT[1], HOG[2], LBP[3] for feature extraction, SVM [4], Bayesian classifier, Random Forest [5] for classification), and different choices may lead to different performance. Obviously, obtaining optimal combination algorithm is complex and time consuming. Moreover, since the representation capability of man-made features is limited, we can't achieve satisfied classification results in most cases.

In recent years, object classification capabilities have dramatically improved due to advances in deep learning and convolutional neural networks (CNNs) have proven to be effective architectures for tackling the complex visual tasks. The model of AlexNet proposed by Krizhevsky et al.[7], has enjoyed a great success in large-scale image recognition and won the 2012 ILSRVC (ImageNet Large-Scale Visual Recognition Challenge) competition with 10.9 percentage points over the second place. In 2014, GoogleNet [8] and VGG [9] won the first and second place in the Imagenet competition. After that, Batch Normalization (BN) proposed by Sergey et al. [10], effectively alleviated the gradient disappearance. ResNet [11] and DenseNet [12] solved the degradation of model performance through skip connection. Compared with the previous network performance, the accuracy and speed both improved greatly. After that, ResNeXt proposed by Xie et al.[13], not only reduces the number of parameters and further improves performance. More recently, Google proposed a novel model of EfficientNet [14] obtained using network architecture search(NAS) technique, boost the performance a lot.

With the rapid development of deep learning techniques in recent years, a variety of deep learning based algorithms for garbage classification have been actively explored. In 2013, Razali et al. [15] proposed a garbage image classification method based on the Elman neural network architecture. In 2016, Sakr et al. [16] presented a garbage image classification algorithm based on the AlexNet framework. In 2017, Yang et al. [17] compared the performance of two classification methods (i.e., SVM and CNN based model) using the TrashNet dataset which contains six categories of garbage (i.e., glass, paper, cardboard, plastic, metal, other). Their CNN based model only achieved 63% accuracy on TrashNet. Later, in 2018, Bircanoğlu et al. [18] designed the RecycleNet algorithm, which further pushes the validation accuracy to 81% on TrashNet. In 2019, Ozkaya et al. [19]achieved state-of-the-art 97.86% accuracy on TrashNet using the combined framework of GoogleNet and SVM.

Despite the existing contributions, garbage classification is still an unsolved challenging problem, especially for shape-shifting garbage classification. Lack of large-scale garbage image datasets is a key factor affecting the research of garbage classification algorithms. At present, the largest public garbage dataset is TrashNet which was constructed by Yang et al. [17] with Stanford University. However, the total number of images in TrashNet is only 2527 including six categories, i.e., glass, paper, cardboard, plastic, metal and other trash. Obviously, it can't meet the requirements of large-scale training and testing data for design state-of-the-art deep learning based garbage classification algorithms. In order to fill this gap, we built a new public garbage dataset which incorporates four major categories, 10 sub-categories and 10624 images.

In this paper, a CNN based multi-task garbage classification architecture, called PublicGarbageNet, is proposed. One task is to classify the four major categories, and the other task is to identify 10 sub-categories. The two classification tasks are related to each other, and the joint loss function is helpful to improve the accuracy of garbage recognition. In addition, after using several effective tricks such as data augmentation, learning rate optimized adjustment, and label smoothing, our PublicGarbageNet achieves state-of-the-art 96.35% accuracy on the new large-scale public garbage dataset.

The reminder of this paper is organized as follows. In Section 2, we introduce how to construct a new public garbage dataset. Then, Section 3 gives a brief introduction to several state-of-the-art CNN architectures for image classification and some useful tricks of network training. Section 4 provides quantitative comparison results of various optimization strategies, and followed by conclusions in Section 5.

2 Public garbage dataset construction

Although the Stanford University's TrashNet dataset can be used to evaluate the performance of garbage classification algorithms, the number of images in the dataset is small, making it difficult to obtain objective results. In view of the abovementioned problem, we construct a new large-scale public garbage dataset. Fig.1 shows the garbage image collection system.

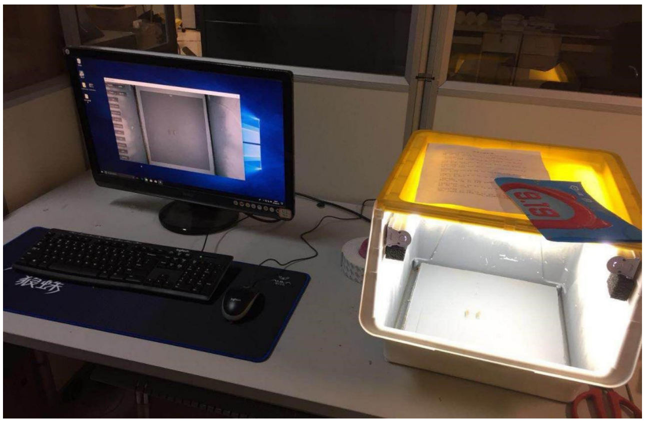

Fig. 1: The garbage image collection system

We utilizes the Unity as development environment to develop the software of image acquisition which supports Windows, MacOS, and Linux systems. The Unity uses the C# programming language, in which the WebCamTexture class can be used to invoke the external camera and display the results through GUI control. In addition, category selection buttons are added to save different kinds of waste images. Each garbage is placed on a gray board as background with a camera mounted on top. The camera captures garbage images with resolution 960×900. In the collection procedure, we mimic the realistic scenario. Specifically, each garbage is freely placed onto the board and taken several photos from different views. In addition, for the wastes such as metal cans, beverage bottles and paper boxes, we squeeze and twist them in different ways.

In total, we collect 10624 images for classification algorithm evaluation. In the new dataset, the garbage images are divided into 10 sub-categories, i.e., kitchen waste, recyclable plastics, unrecyclable plastics, recyclable paper, unrecyclable paper, metal, electronics, glass, textile and hazardous material. The major category of the recyclable garbage contains five sub-categories, i.e., recyclable plastics, recyclable paper, metal, electronic product and glass. The major category of other waste includes three sub-categories, i.e., unrecyclable plastics, unrecyclable paper and textile. Therefore, the number of major categories is four. The detailed statistic information about the new dataset is shown in the Table 1.

3 Methodology

3.1 Network architecture

In 2015, He et al. [11] introduced a deep residual learning framework, a.k.a ResNet, and won the 1st place on the ILSVRC 2015 classification competition. This model adds the skip connections to create a residual mapping, which alleviates the vanishing-gradient problem in training the substantially deeper networks. In 2017 , Huang et al.[12] proposed the dense convolutional network, a.k.a, DenseNet, which connects each layer to all subsequent layers in a feed-forward manner. The links between the features in different channels encourage feature reuse, so that the number of parameters is substantially reduced without accuracy decreasing. In the same year, Xie et al.[13] presented a highly modularized convolutional network architecture, a.k.a ResNeXt, in which a set of transformations with the same topology is aggregated in each block. Empirical evidences show that increasing cardinality (the size of the set of transformations) is able to improve the performance of classification accuracy. In 2019, the EfficientNet proposed by Tan et al. [14] utilizes a simple yet effective component coefficient to uniformly scale three dimensions of network width, depth and resolution, which further reduces the network parameters and improves the recognition accuracy.

At present, most cities in China adopt the "four classification" standard for garbage classification, that is, kitchen waste, hazardous waste, recyclable materials and other waste. In order to implement fine-grained garbage sorting, some major categories are further divided into several sub-categories. For example, The major category of recyclable materials is divided into five sub-categories, i.e., recyclable plastics, recyclable paper, metal, electronic product and glass. Considering the specific scenarios for pubic garbage classification, a CNN based two-task classification algorithm, called PublicGarbageNet, is proposed. This model can not only identifies four major categories of garbage, but also classifies 10 sub-categories of waste. The network architecture of the PublicGarbageNet is shown in Fig.3.

The original garbage image is first fed into the network, and then hierarchical features are extracted using the optimized backbone. Finally, the feature maps are input into two classification branches: one for major category classification, the other for sub-category classification. The two classification tasks are related to each other, and the joint loss function is helpful to improve the accuracy of garbage recognition. We set weighted penalty factors α and fuse the cross entropy losses of two branches as follows:

Table 1: The statistic information of the garbage dataset

Major categories	Sub-categories	Sub-categories number	Major categories number
Kitchen Waste	Kitchen Waste	246	246
Recyclable	Recyclable Plastics	2606	6450
	Recyclable Paper	1803	
	Electronic Product	147	
	Metal	1599	
	Glass	295	
Other Waste	Unrecyclable Paper	1081	3349
	Unrecyclable Plastics	2077	
	textile	191	
Hazardous Waste	Hazardous Material	179	179

Fig. 2: The example images of each sub-category

$$l_loss = \alpha * l_{sub\_loss} + (1 - \alpha) * l_{main\_loss} \quad (1)$$

Among them, l_{sub_loss} is the cross entropy loss of sub-categories classification, l_{main_loss} denotes the cross entropy loss of major categories classification. We set $\alpha < 0.5$ to increase the punishment for major categories recognition errors.

3.2 Data augmentation

For the deep convolution neural network, although the strategies of Batch Normalization and residual skip connection can reduce the gradient varnishing of the network and make the network convergence, deep convolution neural network still has serious overfitting problem when the size of the dataset is small.

Data augmentation is an effective way to expand the dataset and make the dataset as diverse as possible. Using data augmentation can effectively reduce the degree of network overfitting and make the training model have stronger generalization ability.

A series of image augmentation operations are used in the implementation of PublicGarbageNet. The image is flipped randomly horizontally or vertically before the image is input into the neural network. Then we will rotate randomly at a certain angle, and shift the center of the image randomly at a certain distance. Finally, the center clipping operation is carried out to erase the useless pixel of the edge part of the image.

Random erasure is used to erase the clipped image. We randomly select a part of the image, and then replace the pixels in this area with a random Gaussian regularization value. Random erasure is a more effective way to reduce the overfitting degree of PublicGarbageNet for a specific dataset and improve the accuracy of the model.

3.3 Learning rate optimization

In the training process of convolutional neural network, optimizer selection is a crucial factor to reach a better minimum for the algorithm. There are many popular optimizers such as SGD [20], AdaGrad [21], Adam [22]and so on. Among them, Adam is widely used in the actual deep learning network training process. It can realize step annealing adaptively and adjust the network learning rate dynamically.

In the initial training process, the gradients of the network are usually very large. If the learning rate is set too large, it will easily lead to the gradient explosion problem. Therefore, during the initial training, we must set the learning rate to a small value. After a certain training time, the learning rate can be adjusted larger appropriately. At the end period of network training, the learning rate needs to be relatively reduced to a little value, so that the network will achieve better convergence. Loshchilov et al. [23] proposed a dynamic learning rate with warmup which is a dynamic learning rate adjustment strategy.

In this paper, motivated by the idea of warmup learning rate adjustment strategy, the combination of cosine learning rate based on warmup with Adam optimizer is used to training PublicGarbageNet, so that the network can converge to a

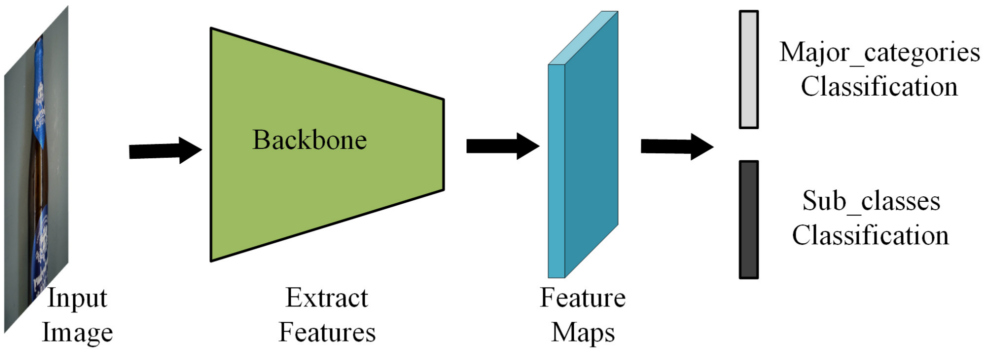

Fig. 3: The network architecture of PublicGarbageNet

better minimum value.

In the initial stage of network training (the process of warmup), the learning rate increases linearly from zero to the initial learning rate.

$$lr = s \times \frac{lr_{init}}{s1} \tag{2}$$

$$s1 = w \times \frac{n}{b} \tag{3}$$

Among them, lr is the adjusted learning rate, lr_{init} denotes the initial learning rate, s is the global step in the training process, $s1$ denotes the total number of warmup training steps, b is the total number of warmup epoches, n represents the total number of training dataset images, and b is size of batch.

After the learning rate reaches the initial learning rate lr_{init}, train h steps according to the initial learning rate, and then gradually reduce the learning rate according to the following Equ.(4).

$$lr = 0.5 \times lr_{init} \times (1 + cosine(\pi \times \frac{s - s1 - h}{S - s1 - h})) \tag{4}$$

Where S is the maximum number of training steps.

3.4 Label smoothing

In supervised learning, the labeling information of data is often not completely correct. The mislabeled sample data will have a certain negative impact on the recognition accuracy. Using the strategy of label smoothing to give the label information a certain fault tolerance rate can effectively improve the generalization ability of the network and improve the accuracy of the test.

The basic idea of the label smoothing strategy is to reduce the dependence of the network on the label, which is a regularization strategy. The basic way to realize this strategy is to adjust the label according to certain rules for an input image, Finally, the adjusted label of one image is the original label with ϵ Probability and is other labels of $1 - \epsilon$ probability. We can reduce the dependence on labels and the impact of wrong labels. For multi-category clasification, the label smoothing is performed as follows:

$$y1 = (1 - \epsilon) \times y + \frac{\epsilon}{c} \tag{5}$$

Where $y1$ denotes the smoothing label, y is the label of the original dataset, c represents the dataset categories, and ϵ is the smoothing factor.

In order to improve the accuracy of PublicGarbageNet, the label smoothing strategy is used in the training process to reduce the impact of the mislabeled samples.

4 Experiments

4.1 Comparisons with different backbones

First of all, four different CNN architectures, namely DenseNet[12], ResNet[11], ResNeXt[13] and EfficientNet [14], are selected as the candidate backbone respectively to test the accuracy and speed of the classification model. Specific network architectures include DenseNet121, DenseNet169, ResNet50, ResNet101, ResNeXt50, ResNeXt101, Efficientnet-B3 and Efficientnet-B4. Finally, we trade off the two factors, i.e., the precision and inference speed and select one backbone suitable for public garbage classification. The training and testing experiments are performed using the Keras library (version 2.21) with TensorFlow backend and the hardware platform of this research is Intel ® core ™ i7-5930k CPU @ 3.50GHz with dual NVIDIA GeForce GTX 1080 Ti GPUs.

More specifically, all input images are reshaped to 224×224 for the backbones of DenseNet, ResNet and ResNeXt, 300×300 for the backbone of Efficientnet-B3, and 380×380 for Efficientnet-B4. The public garbage dataset is divided into training set and test set with the proportion of 5:1. We initialize the parameters of each backbone which was pre-trained on the ImageNet. We also use several data augmentation operations such as random flip, random rotation, random translation, center clipping, random erasure, to expand the dataset and make the dataset as diverse as possible. In the training process, the batch sizes of DenseNet, ResNet and ResNeXt are set to 32 and 16 for the EfficientNet. The initial learning rate is set to 0.0001, and then adjusted according to the Equ.(4). The epoch of warmup w is set to 5, and the h is equal to 5. The penalty factor α of the two-branch modulars is set to 0.33 and the smoothing factor for label smoothing is set to 0.1.

Comparison from the perspective of accuracy, using EfficientNet-B4 achieved the highest rate of 96.76%, followed by EfficientNet-B3, 96.47% and DenseNet169, 96.35%. The worst is the result of ResNet50, only 93,23%. However, due to to the large number of model parameters, the inference time using EfficientNet-B4 and EfficientNet-

Table 2: The comparison results using different backbones

Backbone	Input size	Accuracy	Inference time	Network parameters
DenseNet121	224×224	95.35%	20.8ms	7.05M
DenseNet169	224×224	96.35%	23.7ms	12.66M
ResNet50	224×224	93.23%	21.2ms	23.58M
ResNet101	224×224	93.29%	36.6ms	42.64M
ResNeXt50	224×224	95.94%	21.4ms	23.06M
ResNeXt101	224×224	95.47%	35.8ms	42.28M
EfficientNet-B3	300×300	96.47%	32.3ms	10.79M
EfficientNet-B4	380×380	96.76%	48.0ms	17.69M

Table 3: The results of the ablation study

Algorithm	Multiple branches	Data augmentation	Warm up learning rate	Label smoothing	Accuracy
PublicGarbageNet	Y	Y	Y	Y	96.35%
	N	Y	Y	Y	95.88%
	Y	N	Y	Y	91.23%
	Y	Y	N	Y	93.85%
	Y	Y	Y	N	95.88%

B3 is up to 48ms and 32.3ms. Although the accuracy rate using DenseNet169 is slightly lower than EfficientNet-B3 and EfficientNet-B4, but the inference time is shorter with 23.7ms. Therefore, considering the two factors of accuracy and speed, we choose DenseNet169 as the preferred backbone.

4.2 Ablation study

Apart from analyzing the impact of different backbone selections on the performance of classification networks, we also tested the influence of the other network component and training skills on the performance of the proposed algorithm. In particular, the ablation study including multi-task learning, data augmentation, dynamic learning rate adjustment and label smoothing, was performed on the public garbage dataset.

Multi-task learning

In order to verify the effectiveness of multi-task learning, we replace the last two-branch structure to a single branch structure, that is, only 10 sub-category garbage are classified and keep the training skills such as data augmentation, warmup learning rate adjustment, label smoothing unchanged. The test result is shown in Table 3. The accuracy of two-task network is 0.47% higher than that of single-task network.

Data augmentation

As mentioned above, data augmentation is an effective way to expand the dataset. Using the strategy of data augmentation can effectively reduced the degree of network overfitting problem making the algorithm more robust. We utilized several conventional operations of data augmentation such as random flip, random rotation, random translation, center clipping, random erasure. The result in Table 3 shows that the strategy of data augmentation significantly improves the accuracy of classification. Specifically, the accuracy of the model using data augmentation is 5.12% higher than that of the model without using data augmentation.

Dynamic learning rate adjustment

The learning rate of the Adam is set to a fixed learning rate, 0.0001. The other configuration of the PublicGarbageNet is unchanged. The influence of using warmup cosine learning rate on the accuracy of the algorithm is quantitatively analyzed. The test results in Table 3 confirm that dynamic learning rate adjustment strategy can improve the accuracy 2.5%.

Label smoothing

In order to quantitatively analyze the effectiveness of label smoothing, we remove the label smoothing operation (the smooth factor is set to 0) and keep other parameters in the PublicGarbageNet unchanged. The test results show that the skill of label smoothing has certain beneficial effect on model accuracy.

5 Conclusion

Garbage classification not only helps to improve the quality of waste, which benefits for the end incineration (or landfill) better harmless treatment, but also strengthens the recycling of renewable resources. Despite the existing contribution in literature, classification of real-scenario garbage with high accuracy is still a challenging problem. In this paper, we first built a large-scale public garbage dataset containing four major categories and 10 sub-categories of waste with a total of 10624 samples. Then, the two-task CNN based garbage classification algorithm, called PublicGarbageNet, was proposed. After using a series of useful tricks such as optimal backbone selection, multi-task learning, data augmentation, warmup dynamic learning rate adjustment and label smoothing, the accuracy of the proposed PublicGarbageNet achieves 96.35%.

References

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005.

[3] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[4] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[5] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[13] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1492–1500.

[14] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[15] Z. B. Razali, "Conceptual design of an automated green technology sorting device for mixed household waste," *Journal of Advance of Environmental Biology*, vol. 7, no. 12, pp. 3681–3689, 2013.

[16] G. E. Sakr, M. Mokbel, A. Darwich, M. N. Khneisser, and A. Hadi, "Comparing deep learning and support vector machines for autonomous waste sorting," in *2016 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*. IEEE, 2016, pp. 207–212.

[17] M. Yang and G. Thung, "Classification of trash for recyclability status," *CS229 Project Report*, vol. 2016, 2016.

[18] C. Bircanoğlu, M. Atay, F. Beşer, Ö. Genç, and M. A. Kızrak, "Recyclenet: Intelligent waste sorting using deep neural networks," in *2018 Innovations in Intelligent Systems and Applications (INISTA)*. IEEE, 2018, pp. 1–7.

[19] U. Ozkaya and L. Seyfi, "Fine-tuning models comparisons on garbage classification for recyclability," *arXiv preprint arXiv:1908.04393*, 2019.

[20] L. Bottou, "Stochastic gradient descent tricks," in *Neural networks: Tricks of the trade*. Springer, 2012, pp. 421–436.

[21] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2121–2159, 2011.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[23] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.